

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email editorial.bmjopen@bmj.com

BMJ Open

Reaching Consensus on Reporting Patient and Public Involvement (PPI): Methods and lessons learned from the development of reporting guidelines

Journal:	BMJ Open
Manuscript ID	bmjopen-2017-016948
Article Type:	Research
Date Submitted by the Author:	21-Mar-2017
Complete List of Authors:	<p>Brett, Jo ; Oxford Brookes University, Department of Applied Health and Professional Development</p> <p>Staniszewska, Sophie; University of Warwick, RCN Research Institute</p> <p>Simera, Iveta; University of Oxford, EQUATOR Network, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences</p> <p>Seers, Kate; Royal College of Nursing Research Institute, School of Health & Social Studies, University of Warwick,</p> <p>Mockford, Carole; University of Warwick</p> <p>Goodlad , Susan ; University of Coventry, Centre for Research in psychology, Behaviour and Achievement</p> <p>Altman, Doug; Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences</p> <p>Moher, David; Ottawa Hospital Research Institute, Ottawa Methods Centre</p> <p>Barber, Rosemary; Sheffield University, SchARR, Public Health</p> <p>Denegri, Simon; NIHR, National Director for Public Participation and Engagement in Research; INVOLVE,</p> <p>Entwistle, Andrew; Patient/Carer Reviewer,</p> <p>Littlejohns, Peter; Kings College London, Primary care and public health</p> <p>Morris, Christopher; University of Exeter, University of Exeter Medical School</p> <p>Suleman , Rashida; University of Warwick Warwick Medical School, Patient Partner, UNTRAP</p> <p>Thomas, Victoria; National Institute for Health and Clinical Excellence, Patient and public involvement unit</p> <p>Tysall, Colin; Patient reviewer,</p>
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Patient-centred medicine
Keywords:	Patient and Public Involvement, reporting guideline, methods, delphi

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

GRIPP2 methods paper 130217

**Reaching Consensus on Reporting Patient and Public Involvement (PPI):
Methods and lessons learned from the development of reporting guidelines**

Brett J, Staniszewska S, Simera I, Seers K, Mockford C, Goodlad S, Altman D, Moher D, Barber R, Denegri S, Entwistle A, Littlejohns P, Morris C, Suleman R, Thomas V, Tysall C

Dr Jo Brett, Department of Applied Health and Professional Development, Faculty of Health and Life Sciences, Oxford Brookes University, Jack Straw's Lane, Marston, Oxford OX3 0FL

E-mail: Jbrett@Brookes.ac.uk

Dr Sophie Staniszewska, RCN Research Institute, Warwick Medical School, Division of Health Sciences, University of Warwick, Coventry CV4 7AL. Tel: 024 7615 0622

Dr Iveta Simera, EQUATOR Network, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, OX3 7LD, UK

Prof Kate Seers, Director, RCN Research Institute, Warwick Medical School, Division of Health and Social Care Research, University of Warwick, Coventry CV4 7AL

Dr Carole Mockford, RCN Research Institute, Warwick Medical School, Division of Health Sciences, University of Warwick, Coventry CV4 7AL

Dr Susan Goodlad, Centre for Research in Psychology, Behaviour and Achievement, Coventry University, Coventry, CV1 5FB, UK.

Prof Doug Altman, EQUATOR Network, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, OX3 7LD, UK

GRIPP2 methods paper 130217

Prof David Moher, Ottawa Hospital Research Institute, Ottawa, Canada; School of Epidemiology, Public Health, and Preventive Medicine, University of Ottawa, Ottawa, Canada.

Dr Rosemary Barber, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

Mr Simon Denegri, NIHR National Director for Patients and the Public in Research and Chair, INVOLVE, UCL, UCL School of Life and Medical Sciences, Gower Street, London WC1E 6BT

Dr Andrew Entwistle, UNTRAP, University of Warwick, Coventry CV4 7AL

Prof Peter Littlejohns, MD, FRCP, Faculty of Life Sciences, King's College London, Addison House, Guys Campus, London, SE1 1UL

Dr Christopher Morris, Senior Research Fellow in Child Health, Peninsula Cerebra Research Unit (PenCRU), University of Exeter Medical School, Exeter, UK

Mrs Rashida Suleman, UNTRAP, University of Warwick, Coventry CV4 7AL

Ms Victoria Thomas, Public Involvement Programme, National Institute for Health and Care Excellence (NICE), London, UK

Mr Colin Tysall, UNTRAP, University of Warwick, Coventry CV4 7AL

Keywords: Patient and public involvement, Reporting guidelines, Systematic Review, Delphi survey, Consensus meeting

Lay summary

Patient and public involvement (PPI) in health and social care research has grown over the past decade. However, researchers do not always report the results of their PPI which has led to a poor evidence-base.

Working closely with patients and the public, we have developed guidance for people writing about public involvement to suggest what details to report. We carried out a thorough assessment of studies in this area, and used a Delphi survey to ask 143 people who are knowledgeable about this topic for their opinions about what should be included in the guidance. The Delphi method consists of a series of questionnaires over a specific time period to find out if there is agreement among experts about the topic under discussion. We found strong agreement on a number of issues to include in the guidance from the 112 people who completed three rounds of Delphi questionnaires. We also held a one day meeting to find out if any additional issues where we hadn't reached agreement should be included in the guidance.

This paper describes the methods we used to develop guidelines to encourage better reporting of PPI, and describes challenges and lessons learnt throughout the development process.

ABSTRACT:

Introduction:

Patient and Public Involvement (PPI) is inconsistently reported in health and social care research. Improving the quality of how PPI is reported is critical in developing a higher quality evidence base to gain a better insight into the methods and impact of PPI. This paper describes the methods used to develop and gain consensus on guidelines for reporting PPI in research studies (GRIPP2).

Methods

There were three key stages in the development of GRIPP2: identification of key items for the guideline from systematic review evidence of the impact of PPI on health research and health services; a three phase online Delphi survey with a diverse sample of experts in PPI to gain consensus on included items; and a face to face consensus meeting to finalise and reach definitive agreement on GRIPP2. Challenges and lessons learnt during the development of the reporting guidelines are reported.

Discussion

The process of reaching consensus is vital within the development of guidelines and policy directions, although debate around how best to reach consensus is still needed. This paper discusses the critical stages of consensus development as applied to the development of consensus for GRIPP2, and discusses the benefits and challenges of consensus development.

Keywords: Patient and public involvement, reporting guidance, Consensus, Delphi methods

Strengths and limitations of this study

- This is the first international study to develop reporting guidelines for patient and public involvement in health and social care research
- GRIPP2 has been developed using the Equator network robust methods for the development of reporting guidelines
- Lack of MESH terms for PPI, inconsistency of indexing between databases, large number of titles in searches, and difficulty locating evidence of PPI in the papers lead to time consuming and costly systematic review
- While the online Delphi survey provided a pragmatic and anonymous process for consensus, challenges were encountered with selection bias in the sample, avoiding response fatigue, and decision-making around the presentation of data
- Success of the consensus meeting was due to care planning and the critical role of the facilitator.

BACKGROUND:

Patient and public involvement (PPI) has become an embedded part of health research nationally and internationally, with a rapidly developing evidence base. INVOLVE defines public involvement in research as research being carried out ‘**with**’ or ‘**by**’ members of the public rather than ‘**to**’, ‘**about**’ or ‘**for**’ them. This includes, for example, working with research funders to prioritise research, offering advice as members of a project steering group, commenting on and developing research materials, undertaking interviews with research participants, identifying themes in the data collected, and aiding dissemination through advocacy.

However, a number of reviews have identified the inconsistency of reporting PPI within papers [1-3]. This may be attributable to a range of reasons including weaknesses in the way the studies were conducted, under-valuing the importance of reporting the results of the PPI, or not recognising the importance of contributing to the PPI evidence base. Poorly reported PPI can lead to a weaker understanding of the evidence base of what works, for whom, in what context and why. This weaker understanding means it is more difficult to implement the findings of studies in terms of best PPI practice and enhancing future PPI.

The challenges in relation to inconsistent reporting in health research more generally led to the development of the EQUATOR Network which promotes transparent and accurate reporting of research studies and has enhanced the quality of research reporting through the promotion of guidelines such as CONSORT (for reporting randomised controlled trials) [4] and STROBE (for the reporting of observational studies) [5]. These are now widely used by researchers and journals [6,7]. While the rate of published PPI studies has increased, there has been a lack of guidance for researchers reporting patient and public involvement, which prompted the development of the Guidance for Reporting Patient and Public Involvement (GRIPP) guideline [3]. This paper reports the methods used to develop an updated version of the GRIPP (GRIPP 2) through rigorous systematic reviews and the development of consensus using the method proposed by the EQUATOR Network [8]. The final checklist, structured in a long form (GRIPP-LF) and a short form (GRIPP-SF), is presented in a companion paper [9].

METHODS

Three stages recommended by EQUATOR in the development of reporting guidelines were followed for the development of GRIPP2 [8]: First, a systematic review of the current evidence of the impact of patient and public involvement in international health and social care research and on NHS services in the UK; second, a three phase online Delphi study to gain consensus on the items included in the GRIPP checklist identified from the evidence; and third, a face to face consensus meeting with an expert panel to resolve divergences and any remaining uncertainties following the Delphi study, and to improve the content and clarity of the checklist. This paper describes the methods used and highlights the challenges and lessons learnt from developing GRIPP2.

1) Systematic reviews

GRIPP2 was informed by two systematic reviews [1,2]. Brett and colleagues aimed to assess the international evidence of the impact of PPI on health and social care research and patients, researchers and communities involved (PIRICOM), while Mockford and colleagues aimed to assess the impact of PPI on the National Health Service (NHS) in the UK. The strength of a systematic review lies in the ability to employ a robust and effective search strategy to efficiently integrate existing information and provide data for rational decision making [10]. For each of the systematic reviews an advisory group, including two lay members and other stakeholders with expertise in PPI and systematic reviews, was established to oversee the systematic review process. The advisory groups were consulted at each stage of the review process through regular meetings and by email.

Search strategies combining title and abstract words and data base headings relating to PPI were used to locate the evidence of the impact of PPI.

Searches were conducted by experienced systematic reviewers (JB, CM) in the electronic databases: Medline, CINAHL, PsychINFO, Health Management

Information Consortium, British Nursing Index, Social Science Citation Index, Conference Papers Index, the Cochrane Library, Embase and Web of Science. Hand searching of reference lists of papers and of key journals was also conducted [11]. Grey literature (unpublished reports) was identified through searches in InvoNet, NHS Evidence, the Kings Fund Library, National Library for Health, Joseph Rowntree Fund, and obtained by contact with experts in the field [12-16].

Title and abstract searches were conducted to narrow down the number of papers ordered. A set of inclusion and exclusion criteria was used to select papers for the review. The reliability of the decision process is increased if all papers are independently assessed by more than one researcher, and the decisions shown to be reproducible. A large number of titles and abstracts obtained from the searches, with a high level of consistency in 10% of titles and abstracts viewed by two researchers. A pragmatic decision was therefore taken for one researcher in each of the systematic reviews to view the rest of the titles and abstracts.

The papers obtained were checked against the inclusion and exclusion criteria ([1,2], and then quality assessed using the Critical Appraisal Skills Programme [17]. Grey literature was assessed using the checklist developed by Dixon-Woods (2005) as used by Hubbard et al (2007) to review grey literature from cancer studies [18,19]. Descriptive tables were developed to summarise the evidence.

In order to identify the items for the original GRIPP guideline from the evidence reported, the research team carefully considered each issue in relation to several criteria: (i) whether the information was important to report within a paper that included some level of PPI, (ii) whether it would contribute to enhancing the evidence-base of PPI reporting more generally, and (iii) where the information should be reported to create greater transparency and so enhance the ease of future synthesis [3]. Criteria were used to consider where an aspect of PPI should be reported within the structure of a paper, which helped to structure the guideline according to the key sections usually expected within a paper. This process was repeated with updated literature gained for GRIPP2. The aim was to create a guideline that was logically structured and could be easily used by authors when writing PPI papers and reports, editors and peer reviewers when reviewing

manuscripts for publication, and also for readers in critically appraising published articles and reports [3].

2. Delphi survey

The Delphi technique was used to gain consensus on the updated reporting guideline. The Delphi process sought expert opinion on the included items in the guideline checklist, through several rounds of feedback and revisions to develop consensus. As the evidence base from which the reporting guideline items were identified was suboptimal in terms of consistency of reporting, this step was essential to harness subjective judgements from key stakeholder groups in a systematic way, and to comment on the suitability and comprehensiveness of the items selected. Three Delphi rounds were selected to gain consensus on the GRIPP2 guideline, reflecting previous EQUATOR guidance development methods.

Ethical approval for the study was secured by the Centre for Educational and Industry (CEI) at the University of Warwick who gained generic approval from the University Ethics Committee for all of its online surveys. The committee reviewed the rigorous survey procedures CEI had in place and granted generic ethical approval for its robust process and procedures. The GRIPP2 Delphi survey was assessed by CEI as being covered by the generic approval.

Identification of experts for the Delphi panel was based on the following criteria: individuals with knowledge and/or experience of PPI; individuals working in the field identified through key networks such as INVOLVE; individuals identified through key PPI citations. Individuals were also identified by using a snowball recruitment method where participants forwarded information about the study to other eligible organisations and individuals. This resulted in a diverse sample of stakeholders, including academics, health professional individuals and organisations, patients, carers, patient charities, patient support groups, funders, editors of health related journals, international organisation networks such as the Health Technology International (HTAi) Citizen and Patient Involvement Group, and other European representatives. Our final sample of panellists agreeing to participate in the Delphi process was 143. This composed of 56 researchers, 42 patient and carers, 18

charities or support groups, 9 members of INVOLVE, the UK National Health Service (NHS) Advisory group for public involvement in research, 5 representatives of the UK National Institute for Health Research (NIHR), and two Editors of Health and Social Care Research journals. Eleven of the Delphi participants were from international PPI networks. While there are no guidelines that recommended the number of participants that should be included in Delphi panels [20], this was sufficiently large to gain diverse views, while providing a manageable amount of data to analyse between the Delphi rounds. This sample size is similar to those involved in the development of previous EQUATOR guidelines [8, 21].

An electronic online survey was chosen as a practical form of administration. An electronic survey programme called Snap Survey [22] was used which offered diverse and flexible function options which were designed and tested in-house by the Centre for Education and Industry, Warwick University. A pilot study with ten participants was conducted to check comprehension and acceptability of the questionnaire, and these participants were not involved in the final Delphi survey. A telephone helpline was available for those who were not able to complete the questionnaire online, and in three cases a paper questionnaire was sent and completed by participants and manually entered by the researcher

The Delphi study was conducted over five months (September 2013 to January 2014). Each round of the Delphi process was open for four weeks, with a two to three week turn around for researchers between each stage. Detailed information about the project was sent to participants two weeks before the Delphi-survey started, to ensure respondents had informed decision making around participating and to optimise response rates in each round of the Delphi. Consent was gained prior to participation in the survey. Reminder emails were sent two weeks after the start of each Delphi survey if experts had not responded. Unique identifiers were used to enable personalised emails containing a survey link to be sent to participants to aid survey administration and to allow monitoring of responses and issue of timely reminders to non-respondents. Confidentiality was maintained and all questionnaires were only identifiable by a code, with all data kept on a file protected computer system. Only amalgamated results were reported rather than individual responses.

i) Round 1

In round 1, participants were asked to rate each of the checklist items from 1 to 10, or no judgement. A rating of 1 meant that the respondent considered the item to be unimportant and should be dropped from checklist. A rating of 10 meant that the respondent considered the item to be very important and must be included. Each point on the scale had a descriptor. Space was provided against each item for free text comments to suggest refinements, reiterations or to suggest additional items.

One hundred and forty-three experts took part in this first round. Two researchers analysed the results of round 1 and free text comments were examined to inform any potential additional items.

ii) Round 2

In round 2, participants were asked to rate the GRIPP items again, including any additional items suggested in free text data in round 1. For each item, panellists were given their previous rating, and were also presented with group summary ratings (medians, IQRs), along with all anonymised free text comments from round 1.

The panellists were asked to re-rate the items and add further comments, if desired. Panellists were informed that following round 2, consensus would be defined by the consistency of median scores between rounds (median ≥ 8 = high importance, median 6 or 7 = moderate importance, and median ≤ 5 = low importance), and the absence of significant issues noted in the text comments. If items scored a median score of ≥ 8 in round 1 and round 2, with low inter quartile ranges (IQRs), they had reached consensus for inclusion.

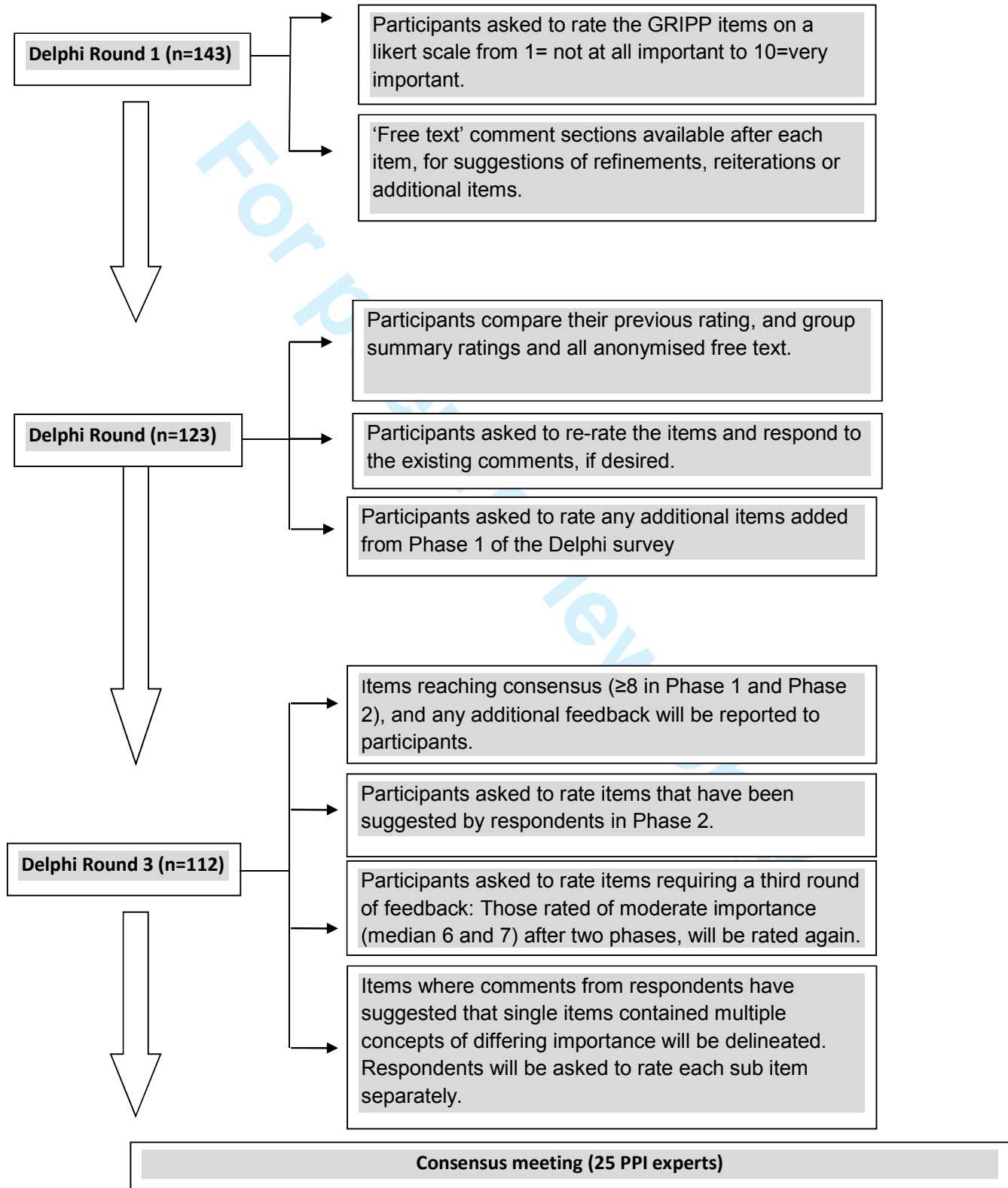
One hundred and twenty three experts took part in the second round. Two researchers analysed the results of round 2 and agreed items to be included in the third round.

iii) Round 3

For round 3, the results of items reaching consensus (from rounds 1 and 2), and any additional feedback were presented. Round 3 also included new items introduced in Round 2, and items rated of moderate importance (median score 6 or 7) in either or both round 1 and round 2). Additionally, phase 3 included items where comments suggested that single items contained multiple concepts of differing importance. For these, concepts were delineated and respondents were asked to rate each sub item separately. Items scoring ≤ 5 in both round 1 and round 2 were excluded.

Figure 1: The Consensus Delphi Exercise

The Delphi process is summarised in figure 1.



3. The face-to-face consensus meeting

The final stage of the project was a 1 day consensus meeting with 25 experts representing a range of key stakeholders, including lay representatives (n=8), health and social care organisations (n=6), journal editors (n=2), and academics working in the field of PPI (n=9). The agenda for the day included scene-setting for this event through presentations on relevant background topics, including details of the systematic review evidence and results of the Delphi exercise. Materials were sent to participants 2 weeks before the meeting, including: the agenda, participant list, one or two key papers, and the results of the Delphi exercise. Consent was gained on the day of the meeting.

The face-to face consensus meeting followed an approach similar to the Nominal Group Technique [23,24] by using small group discussion, sharing of ideas and voting techniques. The detailed discussions at the meeting focussed on those items which only reached moderate consensus (n=7). Participants were divided into 4 roundtable groups, with a diverse group of stakeholders in each group. Each group had 20 minutes to discuss each of the items. Participants were encouraged to voice their opinions, with a prerequisite that ‘all were equal and every contribution is valid’ [25]. Opinions arising during their conversations were captured through the use of different media, for example colour coded cards, post-it notes, and large pieces of paper placed on the tables. This method intended to enhance creative thinking, expression and communication [25,26]. While a professional facilitator hosted the meeting, each table also nominated a ‘table host’ whose role it was to keep a focus and to encourage all participants to contribute to the discussion.

After the discussion, each table was asked to feedback their comments on each item. After all tables had fed back, each table had a further 5 minutes to discuss their decision on whether or not to include the item, and one vote from each table was recorded. Consensus on whether to include or exclude the item was achieved if 3 or more tables were in agreement. If two tables voted to include, and two tables voted to exclude, further discussion as a group was conducted, with the facilitator recording each viewpoint on a flip chart using the words spoken by the participants. Consensus was then gained through individual votes.

The second half of the consensus meeting discussed issues arising in respect of content and face validity of the checklist. Participants were asked to check the wording and make any suggested changes to wording directly onto each item. Comments around the comprehensibility of the item were also sought. Suggested modifications were made by the research team following the consensus meeting.

A key session towards the end of the face-to-face meeting was held to discuss the 'knowledge translation' strategy to assist with the translation of the reporting guideline into practice. A publication strategy was also developed, and discussions included how the guideline can be implemented by journals.

DISCUSSION AND CONCLUSION:

PPI has the potential to improve the quality, relevance and impact of health research, whilst also improving the transparency of the process and the accountability to the wider community of the researchers themselves. However poor reporting of PPI in research studies lead to a range of difficulties. If readers are not given enough detail about how PPI was conducted in a study, they are left with an incomplete picture which makes its appraisal very difficult and has implications for the judgment of reliability and the interpretation of results [6,27].

There are moral and ethical imperatives for the good reporting of research which cannot be ignored. This has been recognised through the EQUATOR Network, which has greatly improved the reliability and value of medical research by promoting transparent and accurate reporting of research studies through guidelines such as CONSORT for RCTs and PRISMA for systematic reviews [27,28]. This paper has detailed the development of consensus on the items that researchers should consider when reporting PPI. The lessons learned are described below.

Important aspects and lessons learned from the development process

i) Systematic Reviews

Systematic review evidence on the impact of PPI on health and social care research and on the National Health Service (NHS) was used to underpin the items included in the original GRIPP and updated for the GRIPP2[1,2]. While the evidence identified in the systematic review was sufficient to identify key areas of importance to structure specific criteria for the reporting guidelines, several pitfalls can be highlighted in conducting systematic reviews around PPI.

Searching databases to identify potential papers for these reviews provided a number of challenges. As there are no Medical Subject Heading (MeSH) for ‘PPI’ a combination of search terms was used for the electronic databases. A lack of specific search terms led to a lack of sensitivity in the initial searches, resulting in a large number of papers identified initially [1,2]. Databases are not consistent in their indexing of studies relating to PPI which poses many challenges for developing search strategies that aim to locate these papers. Databases also vary the search terms used, which means the search strategies need to vary by database, increasing the complexity of searching and the potential for error.

This phase of the study was therefore cumbersome and costly. Standardising the terms used for PPI would not only improve search strategies for future reviews of PPI evidence, but also improve comprehension around PPI among health researchers. Standardised terms could then be adopted by electronic databases improving the MeSH search terms. Care was taken with decisions about where and how to search. For example, a dearth of peer-reviewed evidence on PPI studies indicated the importance of searching the grey literature. Restricting the searches to electronic databases, which consist mainly of references to published peer-reviewed journal articles, could have excluded many PPI papers, leading to publication bias.

While it is recommended that all abstracts are reviewed by two researchers in systematic review methodology [11], this greatly increases the cost and time of the process [29]. Evidence shows only an 8% improvement in identification of relevant papers when all abstracts are reviewed by two researchers. A large number of

abstracts were retrieved in the searches, therefore for pragmatic reasons, 10% of abstracts were reviewed by two researchers. A high level of consistency was reported, indicating the reliability of the searches.

The quality of the evidence was also very difficult to assess as the PPI evidence was often inconsistently reported and difficult to locate. For example, PPI studies were sometimes reported in full in either the methods or the discussion of the paper [1,2]. Quality assessments were therefore conducted to assess inclusion, but not used to weigh the papers. If studies were fatally flawed in terms of their quality, they were excluded. Care should always be taken when interpreting the results of critical appraisals as they can be biased because of the subjective nature in which decisions are made by researchers. One study compared CASP with the Quality Framework (QF) and intuitive judgement by expert opinion [30] and found that no difference could be detected between the different critical appraisal tools and intuitive judgement by expert opinion.

ii) Delphi Survey

Following the update of the original GRIPP checklist, a three-phase online Delphi survey was conducted to gain consensus on all items included in the checklist. Other methods were considered including nominal groups [31], analytic hierarchy process (AHP) technique [32] and use of separate working groups gathering consensus through focus groups [33].

From a pragmatic view point, Delphi methods allowed a large number of geographically dispersed stakeholders to be involved in the consensus process which may not have been possible through alternative consensus processes due to time and cost limitations [34]. Respondents could complete the questionnaire at their leisure and this reduced time pressures and may have allowed for more reflection and contemplation of responses [35].

Additionally, the advantage of this method is that members remain anonymous in responding to individual questions and this is likely to encourage opinions that are free of influences from others and more likely to be 'true' [36,37]. It also provides an iterative process with controlled feedback, and average score responses from a group of experts,

providing stability of responses. Furthermore, it recognises and acknowledges the contribution of each participant. This method can therefore facilitate consensus where there is contradictory or insufficient information to make effective decisions [35-38].

The disadvantages of Delphi methods, which are perhaps true for any consensus method, are that there is a purposeful selection of ‘experts’ chosen because the respondents’ reputations are known to the researcher. These experts meet a minimal number of criteria of familiarity with the research field, self-rating their expertise [39]. Furthermore, the self-selected sample may be biased in that they are willing to take part and therefore either more favourable and more inclined to agree with items included in the GRIPP2 checklist, or disapproving and more likely to disagree with the items included [39].

Delphi methods traditionally use two or more rounds to gain consensus [40]. The number of rounds depends on the level of initial consensus gained, but may also be controlled by time and cost limitations. Studies focussing on the number of rounds needed in a Delphi survey to achieve consensus suggest that most changes occur in the transition from the first to the second round [41]. When the number of rounds exceeds four, the response rates can be very low due to the response burden on participants [41].

The possibility that participants may alter their estimates in order to conform to the group (conformance), without actually changing their opinions (consensus), was considered. However, evidence suggests that the influence of expert knowledge helps move towards consensus rather than conformance with the ‘median score moving towards true value’ [42].

An electronic survey was chosen over postal surveys. This allowed a fast deployment of surveys and a quick return time. It was relatively low cost, removing the cost and time of printing, posting and data entry. It also saved the participants from the inconvenience of posting a survey back to us. The disadvantage of this method is that the participants had to have access to a computer and be computer literate. This may have biased the sample, particularly excluding some hard to reach patients and carers.

An online survey software, Snap Chat, was adapted by in-house expertise to record the survey responses [21]. Other electronic packages were considered such as Survey Monkey and Zoomerang, GoogleForms, and SurveyGizmo [43-46], but Snap Chat was internally available and proved a reliable and flexible survey software for the study's requirements. Furthermore, using in-house expertise allowed us to tailor and customise the online form to our needs more effectively thus allowing us to refine the design and functionality after piloting. Survey responses were monitored and the survey data were cleaned and analysed to produce the results in the required format.

The electronic software offered a save and return function within the survey, allowing participants to reflect and return to the survey. This also would have minimised the risk of incomplete responses where participants sometimes underestimate the time the survey will take and then run out of time to complete the questionnaire in full.

Consideration was given to the number of categories in the Likert scale for rating the items within the GRIPP2 Delphi survey. Streiner and Norman (2008) have argued that the benefit from large numbers of options is subject to the law of diminishing returns [47], and that from the 7-item scale and upwards, the scales become too cumbersome to use. Any additional benefits are cancelled out by respondent fatigue and reliability plummets [47]. However, in reality we are often asked to rate issues on a scale of 1 to 10, it provides a better opportunity to detect and discriminate when responses are skewed at one end of the scale, and it felt more 'natural' to patient advisors on the research team [48]. This is also how previous Delphi surveys had been scored in the development of reporting guidelines through EQUATOR.

The pilot study highlighted a number of issues of importance, such as providing clear instructions of how to access, complete and submit the questionnaire, and providing a clear deadline for submission. These issues were addressed before the first round in the main study. The pilot study also facilitated a link with the study to a bibliography and to a lay language option on the research webpage.

Different descriptive statistics can be used when feeding back data in each round. While the mean, mode and median scores provide three forms of averages [49], the range, inter-quartile range (IQR) and standard deviation are all measures that indicate the amount of variability within a dataset [49]. As there was the potential for a range of scores for each item, the median was selected as a better score of the average than the mean or the mode. The range is the simplest measure of variability to calculate but can be misleading if the dataset contains extreme values. The IQR reduces this problem by considering the variability within the middle 50% of the dataset. The standard deviation is used to take into account a measure of how every value in the dataset varies from the mean. The interquartile range was therefore used to report the variation from the median scores [50].

The validity and reliability of a Delphi study may be questioned due to the subjective nature of the feedback and potential instability of responses. However, this approach engages a wide range of expertise more effectively than any other group consensus method, and provides a ‘fair’ representation of the views of each participant because each participant has an equal opportunity to have their views taken into account. Furthermore, Delphi methods clearly state the rationale for inclusion or exclusion in the final checklist, whereas other methods of gaining consensus rarely provide such a transparent decision trail for each item. The quality of the Delphi survey was further increased by the quick turnaround between Delphi rounds [51]. The Delphi study for the reporting guidelines provided opinions from a representative sample of all PPI stakeholders, carefully and rigorously collated over 3 rounds of voting. In this study the Delphi methods used brought agreement from a diverse group of stakeholders whose commitment to the project resulted in good response rates at each stage of the Delphi project. The validity and reliability of the process was therefore deemed satisfactory.

Finally, ethical issues were also examined in this Delphi survey. The main ethical issues related to consent to participate, privacy, and confidentiality of the data provided, all of which were addressed in the methods used.

iii) Consensus Face to Face Meeting

The consensus face-to face meeting after the Delphi study finalised the GRIPP2 checklist through informal nominal group techniques for items that had not reached consensus in the Delphi study. An informal approach to voting was adopted where consensus was gained through the round robin process with a gradual move towards synthesis and building consensus rather than through anonymous voting [51]. This approach encourages the sharing and discussion of the reasons for the choices made by each group member, thereby identifying common ground, and a plurality of ideas and approaches [51]. The meeting also ensured face and content validity through small group discussion.

The success of the consensus group was enhanced by careful planning and commitment of the stakeholders. Furthermore, a critical role in the success of the consensus group was that of the facilitator. It is crucial that members of the group understand their roles, take ownership of the process of consensus, and that all feel listened to. A highly skilled facilitator was used to mediate the group process, and to ensure key and timely contributions from all members. This important role helped to ensure that balanced views were recorded through individual and group work.

An important aspect of the consensus meeting was also agreeing a plan of action for dissemination, implementation and adoption of GRIPP2. Evaluation of GRIPP2 will be conducted through ongoing feedback from authors.

Table 1 below summarises the lessons learnt during this study:

Table 1: Summary of lessons learnt:

Systematic Review	<ul style="list-style-type: none">• Lack of MESH terms for PPI in databases made searches broad• No consistency in indexing between databases lead to development of separate search strategies for each database• Difficulty identifying PPI data as written up in methods or discussion sections of papers• Pragmatics of all titles reviewed by two researchers led to 10% reviewed twice
Delphi Study	<p><u>Advantages:</u></p> <ul style="list-style-type: none">• Pragmatic consensus approach to gain large, geographically dispersed group• Participants complete the questionnaire at their leisure• Participants remain anonymous so greater openness and honesty of comments <p><u>Challenges:</u></p> <ul style="list-style-type: none">• Selection bias – as self-selected. Tendency to recruit participants who are more favourable or more disapproving.• Excludes those who do not have a computer and those who are not computer literate.• Decision-making on number of Delphi rounds to gain consensus. Need enough to gain consensus but not too many that you have low response rates for subsequent rounds.• Difficulty of choosing off the shelf data collection packages – importance of customising software.• Issues of response fatigue with too many categories in the Likert scale – the higher the number the greater the response fatigue.• Need for clear instructions, bibliography, and other information in lay language.• Use of median score and inter-quartile ranges vs mean (sd) and modes in presentation of data – to overcome extreme values in dataset
Consensus Meeting	<ul style="list-style-type: none">• Need for openness and informal approach to encourage discussion and plurality of ideas.• Critical to have expert facilitator to guide the consensus meeting to ensure key and timely contributions from all members• Importance of planning for dissemination, implementation and adoption of outcomes

Conclusion

In conclusion, this paper details the methods used in the development of EQUATOR recognised guidelines for the reporting of patient and public involvement (PPI) in research (GRIPP2) [9].

GRIPP2 has been developed using the robust methods used in the development of other EQUATOR guidelines such as CONSORT and STROBE. The development process involved identification of relevant evidence through systematic reviews, consensus of included items through an online Delphi survey with PPI experts, and a face to face meeting of PPI experts to finalise consensus.

Systematic reviews informed the development of items for the reporting guideline. However, lack of MESH terms for PPI, inconsistency of indexing between databases, large number of titles in searches, and difficulty locating evidence of PPI in the papers lead to time consuming and costly systematic reviews. An online Delphi survey provided a pragmatic and anonymous process for consensus, although challenges were encountered with selection bias in the sample, avoiding response fatigue and decision-making over the presentation of data. The success of the consensus meeting was due to careful planning and the critical role of the facilitator.

GRIPP2 has been developed to improve the accuracy and consistency of PPI reporting in health research to improve interpretation and better application in future research. As the field of patient and public involvement develops we expect further development of GRIPP2 to refine and update the guidance.

References:

1. Brett J, Staniszewska S, Mockford C, Sandra Herron-Marx S, Hughes J, Tysall C, Suleman R (2014). Mapping the Impact of Patient and Public Involvement (PPI) on Health and Social Care Research: A systematic review. *Health Expectations*. 17(5) pp. 637–650

2. Mockford C, Staniszewska S, Griffiths F, and Herron-Marx S (2011). The impact of patient and public involvement (PPI) on UK NHS healthcare: a systematic review. *International Journal for Quality in Health Care*, (1353-4505)

3. Staniszewska S, Brett J, Mockford C, Barber R (2011). The GRIPP checklist: Strengthening the quality of patient and public involvement reporting in research. *International Journal of Health Technology Assessment in Health Care*, 27 (4): 391-399

4. Schulz KF¹, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010 Mar 23;340:c332. doi: 10.1136/bmj.c332.

5. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007 Oct 20;370(9596):1453-7.

6. Moher D, Altman D, Shultz K, Simera I, Wager E. Guidelines for Reporting Health Research: A User's Manual. *BMJ Books*. Sept 2014

7. Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG (2010). Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med*;8(1):24. PMID: 20420659

8. Moher D, Schulz KF, Simera I, Altman DG (2010). Guidance for developers of health research reporting guidelines. *PLoS Med*; 7(2):e1000217. PMID: 20169112

9. Staniszewska S, Brett J, Simera I, Seers K, Mockford C, Goodlad S, Altman DG, Moher D, Barber R, Denegri S, Entwistle A, Littlejohns P, Morris C, Suleman R, Thomas V, Tysall C (2016) Guidance for reporting patient and public involvement in health research: A Delphi consensus study to develop GRIPP2-SF and GRIPP2-LF. Submitted BMJ/BMJ Open

10. Cook DJ, Mulrow CD, Haynes RB. Systematic Reviews: Synthesis of Best Evidence for Clinical Decisions. *Annals of Internal Medicine*. 1997;126(5):376- 80

11. Centre for Reviews and Dissemination (CRD). Systematic Reviews: Guidance for undertaking reviews in health care. University of York, 2008<https://www.york.ac.uk/>

GRIPP2 methods paper 130217

12. INVOLVE: InvoNET website www.invo.org.uk. Accessed June 2012.
13. NHS Evidence www.evidence.nhs.uk. Accessed June 2012
14. Kings fund Library. www.kingsfund.org.uk/library. Accessed June 2012
15. National library for Health www.library.nhs.uk. Accessed 2012
16. Joseph Rowntree Fund Library . www.jrf.org.uk/Library. Accessed June 2012
17. Critical Appraisal Skills Programme (CASP) 2014. CASP Checklists (URL used)
Oxford. CASP
18. Dixon-Woods M, Kirk D, Agarwal S et al. Vulnerable Groups and Access, to
Health Care: A Critical Interpretative Review. London: National Co-ordinating Centre
for NHS Service, 2005. Dunham, R. B. (1998): Nominal Group Technique: A Users'
Guide. Wisconsin: University of Wisconsin.
19. Hubbard G, Kidd L, Donaghy E, McDonald C, Kearney N. A review of literature
about involving people affected by cancer in research, policy and planning and
practice. Patient Education and Counseling, 2007; 65: 21–33.
20. Williams, P.L. and Webb, C. (1994) 'The Delphi technique: a methodological
discussion', Journal of Advanced Nursing 19 180-186.
21. Tetzlaff J, Moher D, Chan A (2012). Developing a guideline for clinical trial
protocol content: Delphi consensus survey. Trials 13: 176
22. Snap chat www.snapchat.com. Sept. 2011
23. Harvey N, Holmes CA: Nominal group technique: An effective method for
obtaining group consensus. Int J Nurs Pract 2012, 18:188-194.
24. Gallagher M, Hares T, Spencer J, Bradshaw C, Webb I: The nominal group
technique: A research tool for general practice? Fam Pract 1993, 10:76-81.
25. Brown, J. I. D. (2005). The World Cafe. Shaping Our Futures Through
Conversations That Matter (p. 765). Berrett-Koehler Publishers. Inc.
26. Wilkinson, J. E., et al. (2012). A creative approach to the development of an
agenda for knowledge utilization: outputs from the 11th international knowledge
utilization colloquium (KU 11). Worldviews on Evidence Based Nursing, 9(4), 195–
199.

27. Moher D, Schulz KF, Altman DG. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001;357:1191- 1194.

28. Moher D. Guidelines for reporting health care research: advancing the clarity and transparency of scientific reporting. *Can J Anaesth*. 2009;56:96-101

29. Attree P. Growing up in disadvantage: a systematic review of the qualitative evidence. *Child Care Health Dev* 2004;30:679-89.

30. Dixon-Woods, M., Sutton, A.J., **Shaw. R.L.**, Miller, T., Smith, J.A., Young, B., Bonas, S., Booth, A., Jones, D.R. (2007). Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods. *Journal of Health Services Research and Policy*, 12(1), 42-47.

31. Carney, O., McIntosh, J. and Worth, A. (1996) 'The use of the nominal group technique in research with community nurses', *Journal of Advanced Nursing* 23, 5, 1024-1029.

32. Lai, V.S., Wong, B.K. and Cehung, W. (2002) 'Group decision making in a multiple criteria environment: A case using the AHP in software selection', *European Journal of Operational Research* 137 134-144.

33. Morgan, D.L. (1997) *Focus Groups as Qualitative Research*. 2nd Edition. London:

34. Rogers, M.R. and Lopez, E.C. (2002) 'Identifying critical cross-cultural school Sage.

35. Linstone, H.A. and Turoff, M. (eds.) (1975) *The Delphi Method Techniques and method?*, *Nursing Research* 46(2) 116-118.

36. Goodman, C.M. (1987) 'The Delphi technique: a critique', *Journal of Advanced Nursing* 12 729-734.

37. Snyder-Halpern, R. (2002) 'Indicators of organizational readiness for clinical information technology/systems innovation: a Delphi study', *International Journal of Medical Informatics* 63 179-204.

38. Hasson, F., Keeney, S. and McKenna, H. (2000) 'Research guidelines for the Delphi survey technique', *Journal of Advanced Nursing* 32(4) 1008-1015.

GRIPP2 methods paper 130217

39. Keeney S, Hasson F, McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. *Int J Nurs Stud*. 2001 Apr;38(2):195-200.
40. Cuhls K (1998). The Delphi Method. Fraunhofer Institute for Systems and Innovation Research, Germany
41. vanZolingen, S.J. and Klaassen, C.A. (2003) 'Selection processes in a Delphi study about key qualifications in Senior Secondary Vocational Education', *Technological Forecasting and Social Change* 70 317-340.
42. Munier, F. and Rondé P. (2001) 'The role of knowledge codification in the emergence of consensus under uncertainty: empirical analysis and policy implications', *Research Policy* 20 1537-1551 .
43. Survey Monkey (1999) <https://www.surveymonkey.com/> 1999-2016
44. Zoomerang (online web survey) (1999) <http://www.zoomerang.com/> 1999-2016
45. Google forms (Online web survey) www.google.com/forms February 2007
46. Survey Gizmo (2005)<https://www.surveygizmo.com/> 2005-2016
47. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use* 2008
48. Wittnk DR, Bayer LR. (2003). "The Measurement Imperative" *Marketing Research*. 6 (4): 14-22
49. Altman D. *Practical Statistics for Medical Research* (Chapman & Hall). November 22, 1990
50. Murphy, M.K., Black, N.A., Lamping, D.L., McKee, C.M., Sanderson, C.D.B., Askham, J. and Marteau, T. (1998) Consensus development methods, and their use in clinical guideline development: a review *United Kingdom, Health Technology Assessment* 2(3)49
51. Dunham, Randall, B. (1998) 'Organisational Behaviour: The Delphi Technique', University of Wisconsin School of Business, US

Statement of contribution

Staniszewska, Brett, Simera, Seers, Mockford, Goodlad, Morris, Denegri, Barber, Littejohn, Thomas, Suleman, Entwistle and Tysall made substantial contributions to conception and design. Brett, Staniszewska, Simera, Altman, and Moher made substantial contributions to developing the protocol. Staniszewska, Brett, Goodlad made substantial contributions to the acquisition of data, analysis, and interpretation of data. All authors have been involved in drafting the manuscript or revising it critically for important intellectual content; given final approval of the version to be published.

Transparency declaration: the manuscript is an honest, accurate, and transparent account of the methods for the study being reported; no important aspects of the study have been omitted; and there are no discrepancies from the study

Acknowledgments

We are very grateful to everyone who participated in the Delphi exercise and attending the consensus event. GRIPP 2 was funded by the RCN Research Institute, Warwick Medical School, as part of its strategic focus on patient and public involvement. We are grateful for Sally Crowe who facilitated the consensus workshop. Sophie Staniszewska is part funded by CLAHRC-WM. Peter Littlejohns is supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) South London at King’s College Hospital NHS Foundation Trust. The views expressed are those of the author and not necessarily those of the NHS, the NIHR, or the Department of Health.

Compliance with Ethical Standards:

Ethical approval for the study was secured by the Centre for Educational Development, Appraisal and Research (CEDAR) at the University of Warwick who secured permission to carry out the Delphi Survey from the University of Warwick BSREC committee

Ethical approval for the study was secured by the Centre for Educational and Industry. Appraisal and Research (CEI) at the University of Warwick who gained generic approval from the University Ethics Committee for all of its online surveys. The committee reviewed the rigorous survey procedures CEI had in place and

GRIPP2 methods paper 130217

granted generic ethical approval for its robust process and procedures. The GRIPP2 Delphi survey was assessed by CEI as being covered by the generic approval.

Funding: This study was funded by RCNRI, Warwick Medical School, University of Warwick

Data-Sharing: Please contact Dr Jo Brett if additional information is required: Email: Jbrett@brookes.ac.uk

Conflict of Interest: No authors reported conflict of interests

For peer review only

BMJ Open

Reaching Consensus on Reporting Patient and Public Involvement (PPI) in Research: Methods and lessons learned from the development of reporting guidelines

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2017-016948.R1
Article Type:	Research
Date Submitted by the Author:	18-Jul-2017
Complete List of Authors:	<p>Brett, Jo ; Oxford Brookes University, Department of Applied Health and Professional Development</p> <p>Staniszewska, Sophie; University of Warwick, RCN Research Institute</p> <p>Simera, Iveta; University of Oxford, EQUATOR Network, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences</p> <p>Seers, Kate; Royal College of Nursing Research Institute, School of Health & Social Studies, University of Warwick,</p> <p>Mockford, Carole; University of Warwick</p> <p>Goodlad , Susan ; University of Coventry, Centre for Research in psychology, Behaviour and Achievement</p> <p>Altman, Doug; Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences</p> <p>Moher, David; Ottawa Hospital Research Institute, Ottawa Methods Centre</p> <p>Barber, Rosemary; Sheffield University, SchARR, Public Health</p> <p>Denegri, Simon; NIHR, National Director for Public Participation and Engagement in Research; INVOLVE,</p> <p>Entwistle, Andrew; Patient/Carer Reviewer,</p> <p>Littlejohns, Peter; Kings College London, Primary care and public health</p> <p>Morris, Christopher; University of Exeter, University of Exeter Medical School</p> <p>Suleman , Rashida; University of Warwick Warwick Medical School, Patient Partner, UNTRAP</p> <p>Thomas, Victoria; National Institute for Health and Clinical Excellence, Patient and public involvement unit</p> <p>Tysall, Colin; Patient reviewer,</p>
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Patient-centred medicine, Medical education and training, Medical publishing and peer review
Keywords:	Patient and Public Involvement, reporting guideline, methods, delphi

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Reaching Consensus on Reporting Patient and Public Involvement (PPI) in Research: Methods and lessons learned from the development of reporting guidelines

Brett J, Staniszewska S, Simera I, Seers K, Mockford C, Goodlad S, Altman D, Moher D, Barber R, Denegri S, Entwistle A, Littlejohns P, Morris C, Suleman R, Thomas V, Tysall C

Dr Jo Brett, Department of Applied Health and Professional Development, Faculty of Health and Life Sciences, Oxford Brookes University, Jack Straw's Lane, Marston, Oxford OX3 0FL

E-mail: Jbrett@Brookes.ac.uk

Prof Sophie Staniszewska, RCN Research Institute, Warwick Medical School, Division of Health Sciences, University of Warwick, Coventry CV4 7AL. Tel: 024 7615 0622

Dr Iveta Simera, EQUATOR Network, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, OX3 7LD, UK

Prof Kate Seers, Director, RCN Research Institute, Warwick Medical School, Division of Health and Social Care Research, University of Warwick, Coventry CV4 7AL

Dr Carole Mockford, RCN Research Institute, Warwick Medical School, Division of Health Sciences, University of Warwick, Coventry CV4 7AL

Dr Susan Goodlad, Centre for Research in Psychology, Behaviour and Achievement, Coventry University, Coventry, CV1 5FB, UK.

Prof Doug Altman, EQUATOR Network, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, OX3 7LD, UK

Prof David Moher, Ottawa Hospital Research Institute, Ottawa, Canada; School of Epidemiology, Public Health, and Preventive Medicine, University of Ottawa, Ottawa, Canada.

Dr Rosemary Barber, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

Mr Simon Denegri, NIHR National Director for Patients and the Public in Research and Chair, INVOLVE, UCL, UCL School of Life and Medical Sciences, Gower Street, London WC1E 6BT

Dr Andrew Entwistle, UNTRAP, University of Warwick, Coventry CV4 7AL

Prof Peter Littlejohns, MD, FRCP, Faculty of Life Sciences, King's College London, Addison House, Guys Campus, London, SE1 1UL

Dr Christopher Morris, Senior Research Fellow in Child Health, Peninsula Cerebra Research Unit (PenCRU), University of Exeter Medical School, Exeter, UK

Mrs Rashida Suleman, UNTRAP, University of Warwick, Coventry CV4 7AL

Ms Victoria Thomas, Public Involvement Programme, National Institute for Health and Care Excellence (NICE), London, UK

Mr Colin Tysall, UNTRAP, University of Warwick, Coventry CV4 7AL

Keywords: Patient and public involvement, Reporting guidelines, Systematic Review, Delphi survey, Consensus meeting

Lay summary

Patient and public involvement (PPI) in health and social care research has grown over the past decade, with patients and the public working with researchers to help prioritise, plan, design and conduct research studies. However, researchers do not always report the results of their PPI which has led to a poor evidence-base.

Working closely with patients and the public, we have developed guidance for people writing about public involvement to suggest what details to report. We carried out a thorough assessment of studies in this area, and used a Delphi survey to ask 143 people who are knowledgeable about this topic for their opinions about what should be included in the guidance. The Delphi method consists of a series of questionnaires over a specific time period to find out if there is agreement among experts about the topic under discussion. We found strong agreement on a number of issues to include in the guidance from the 112 people who completed three rounds of Delphi questionnaires. We also held a one day meeting to find out if any additional issues where we hadn't reached agreement should be included in the guidance.

This paper describes the methods we used to develop guidelines to encourage better reporting of PPI, and describes challenges and lessons learnt throughout the development process.

ABSTRACT:

Introduction:

Patient and Public Involvement (PPI) is inconsistently reported in health and social care research. Improving the quality of how PPI is reported is critical in developing a higher quality evidence base to gain a better insight into the methods and impact of PPI. This paper describes the methods used to develop and gain consensus on guidelines for reporting PPI in research studies (GRIPP2).

Methods

There were three key stages in the development of GRIPP2: identification of key items for the guideline from systematic review evidence of the impact of PPI on health research and health services; a three phase online Delphi survey with a diverse sample of experts in PPI to gain consensus on included items; and a face to face consensus meeting to finalise and reach definitive agreement on GRIPP2. Challenges and lessons learnt during the development of the reporting guidelines are reported.

Discussion

The process of reaching consensus is vital within the development of guidelines and policy directions, although debate around how best to reach consensus is still needed. This paper discusses the critical stages of consensus development as applied to the development of consensus for GRIPP2, and discusses the benefits and challenges of consensus development.

Keywords: Patient and public involvement, reporting guidance, Consensus, Delphi methods

Strengths and limitations of this study

- This study describes the methods for the development of the first international guidance for the reporting of patient and public involvement in health and social care research
- GRIPP2-LF and GRIPP2-SF has been developed using the Equator network robust methods for the development of reporting guidelines
- Lack of MESH terms for PPI, inconsistency of indexing between databases, large number of titles in searches, and difficulty locating evidence of PPI in the papers lead to time consuming and costly systematic review
- While the online Delphi survey provided a pragmatic and anonymous process for consensus, challenges were encountered with selection bias in the sample, avoiding response fatigue, and decision-making around the presentation of data
- Success of the consensus meeting was due to care planning and the critical role of the facilitator.

BACKGROUND:

Patient and public involvement (PPI) has become an embedded part of health research nationally and internationally, and has the potential to improve the quality, relevance and impact of health research, whilst also improving the transparency of the process and the accountability to the wider community of the researchers themselves. INVOLVE[†] defines public involvement in research as research being carried out ‘with’ or ‘by’ members of the public rather than ‘to’, ‘about’ or ‘for’ them[†]. This includes, for example, working with research funders to prioritise research, offering advice as members of a project steering group, commenting on and developing research materials, undertaking interviews with research participants, identifying themes in the data collected, and aiding dissemination through advocacy.

However, a number of reviews have identified the inconsistency of reporting PPI within papers [1-3]. This may be attributable to a range of reasons including weaknesses in the way the studies were conducted, under-valuing the importance of reporting the results of the PPI, or not recognising the importance of contributing to the PPI evidence base. Poorly reported PPI can lead to a weaker understanding of the evidence base of what works, for whom, in what context and why. This weaker understanding means it is more difficult to implement the findings of studies in terms of best PPI practice and enhancing future PPI.

The challenges in relation to inconsistent reporting in health research more generally led to the development of the EQUATOR Network which promotes transparent and accurate reporting of research studies and has enhanced the quality of research reporting through the promotion of guidelines such as CONSORT (for reporting randomised controlled trials) [4] and STROBE (for the reporting of observational studies) [5]. These are now widely used by researchers and journals [6,7]. While the rate of published PPI studies has increased, there has been a lack of guidance for researchers reporting patient and public involvement, which prompted the development of the Guidance for Reporting Patient and Public Involvement (GRIPP) guideline [3]. While the original GRIPP checklist represented an important starting point in creating high quality PPI reporting, its development drew on systematic review evidence, without broader input from the international PPI

[†]<http://www.invo.org.uk/>
For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

research community. Achieving consensus is now acknowledged as a crucial step in producing a reporting guideline [6]. GRIPP2 addresses this gap by developing consensus within the international PPI community. This paper reports the methods used to develop an updated version of the GRIPP (GRIPP 2) through rigorous systematic reviews and the development of consensus using the method proposed by the EQUATOR Network [8]. The final checklist, structured in a short form (GRIPP-SF) and in a long form (GRIPP-LF), is presented in a companion paper [9]. GRIPP 2-SF is a short checklist for studies where PPI is a secondary or tertiary focus such as in a RCT, and GRIPP 2-LF is a longer checklist for studies where PPI forms the primary focus of a study, such as a paper primarily reporting the impact of PPI on the study. For GRIPP2-LF the entire paper can be shaped by the guidance, with researchers selecting the items of relevance. With GRIPP2-SF researchers could present all the information in a short section or in a separate box.

METHODS

Three stages recommended by EQUATOR in the development of reporting guidelines were followed for the development of GRIPP2 [8]: First, a systematic review of the current evidence of the impact of patient and public involvement in international health and social care research and on NHS services in the UK; second, a three phase online Delphi study to gain consensus on the items included in the GRIPP checklist identified from the evidence; and third, a face to face consensus meeting with an expert panel to resolve divergences and any remaining uncertainties following the Delphi study, and to improve the content and clarity of the checklist. This paper describes the methods used and highlights the challenges and lessons learnt from developing GRIPP2.

1) Systematic reviews

GRIPP2 was informed by two systematic reviews [1,2]. Brett and colleagues aimed to assess the international evidence of the impact of PPI on health and social care research and patients, researchers and communities involved (PIRICOM), while Mockford and colleagues aimed to assess the impact of PPI on the National Health Service (NHS) in the UK. The strength of a systematic review lies in the ability to

employ a robust and effective search strategy to efficiently integrate existing information and provide data for rational decision making [10]. For each of the systematic reviews an advisory group, including two lay members and other stakeholders with expertise in PPI and systematic reviews, was established to oversee the systematic review process. The advisory groups were consulted at each stage of the review process through regular meetings and by email.

Search strategies combining title and abstract words and database headings relating to PPI were used to locate the evidence of the impact of PPI.

Searches were conducted by experienced systematic reviewers (JB, CM) in the electronic databases: Medline, CINAHL, PsychINFO, Health Management Information Consortium, British Nursing Index, Social Science Citation Index, Conference Papers Index, the Cochrane Library, Embase and Web of Science. Hand searching of reference lists of papers and of key journals was also conducted [11]. Grey literature (unpublished reports) was identified through searches in InvoNet, NHS Evidence, the Kings Fund Library, National Library for Health, Joseph Rowntree Fund, and obtained by contact with experts in the field [12-16].

Title and abstract searches were conducted to narrow down the number of papers ordered. A set of inclusion and exclusion criteria was used to select papers for the review. The reliability of the decision process is increased if all papers are independently assessed by more than one researcher, and the decisions shown to be reproducible. However, due to the large number of titles and abstracts obtained from the searches, and with a high level of consistency in 10% of titles and abstracts reviewed by two researchers, a pragmatic decision was taken for one researcher in each of the systematic reviews to view the rest of the titles and abstracts.

The papers obtained were checked against the inclusion and exclusion criteria ([1,2], and then quality assessed using the Critical Appraisal Skills Programme [17]. Grey literature was assessed using the checklist developed by Dixon-Woods (2005) as used by Hubbard et al (2007) to review grey literature from cancer studies [18,19]. Descriptive tables were developed to summarise the evidence.

In order to identify the items for the original GRIPP guideline from the evidence reported, the research team carefully considered each issue in relation to several criteria: (i) whether the information was important to report within a paper that included some level of PPI, (ii) whether it would contribute to enhancing the evidence-base of PPI reporting more generally, and (iii) where the information should be reported to create greater transparency and so enhance the ease of future synthesis [3]. Criteria were used to consider where an aspect of PPI should be reported within the structure of a paper, which helped to structure the guideline according to the key sections usually expected within a paper. This process was repeated with updated literature gained for GRIPP2. The aim was to create a guideline that was logically structured and could be easily used by authors when writing PPI papers and reports, editors and peer reviewers when reviewing manuscripts for publication, and also for readers in critically appraising published articles and reports [3].

2. Delphi survey

The Delphi technique was used to gain consensus on the updated reporting guideline. The Delphi process sought expert opinion on the included items in the guideline checklist, through several rounds of feedback and revisions to develop consensus. As the evidence base from which the reporting guideline items were identified was suboptimal in terms of consistency of reporting, this step was essential to harness subjective judgements from key stakeholder groups in a systematic way, and to comment on the suitability and comprehensiveness of the items selected. Three Delphi rounds were selected to gain consensus on the GRIPP2 guideline, reflecting previous EQUATOR guidance development methods.

Ethical approval for the study was secured by the Centre for Education and Industry (CEI) at the University of Warwick who gained generic approval from the University Ethics Committee for all of its online surveys. The committee reviewed the rigorous survey procedures CEI had in place and granted generic ethical approval for its robust process and procedures. The GRIPP2 Delphi survey was assessed by CEI as being covered by the generic approval.

Identification of experts for the Delphi panel was based on the following criteria: individuals with knowledge and/or experience of PPI; individuals working in the field identified through key networks such as INVOLVE; individuals identified through key PPI citations. Individuals were also identified by using a snowball recruitment method where participants forwarded information about the study to other eligible organisations and individuals. This resulted in a diverse sample of stakeholders, including academics, health professional individuals and organisations, patients, carers, patient charities, patient support groups, funders, editors of health related journals, international organisation networks such as the Health Technology International (HTAi) Citizen and Patient Involvement Group[†], and other European representatives. Our final sample of panellists agreeing to participate in the Delphi process was 143. This composed of 56 researchers, 42 patient and carers, 18 charities or support groups, 9 members of INVOLVE, the UK National Health Service (NHS) Advisory group for public involvement in research, 5 representatives of the UK National Institute for Health Research (NIHR), and two Editors of Health and Social Care Research journals. Eleven of the Delphi participants were from international PPI networks. While there are no guidelines that recommended the number of participants that should be included in Delphi panels [20], this was sufficiently large to gain diverse views, while providing a manageable amount of data to analyse between the Delphi rounds. This sample size is similar to those involved in the development of previous EQUATOR guidelines [8, 21].

An electronic online survey was chosen as a practical form of administration. An electronic survey programme called Snap Survey [22] was used which offered diverse and flexible function options which were designed and tested in-house by the Centre for Education and Industry, Warwick University. A pilot study with ten participants was conducted to check comprehension and acceptability of the questionnaire, and these participants were not involved in the final Delphi survey. The pilot study highlighted a number of issues of importance, such as providing clear instructions of how to access, complete and submit the questionnaire, and providing a clear deadline for submission. These issues were addressed before the first round in the main study. The pilot study also facilitated a link with the study to a bibliography and to a lay language option on the research webpage.

[†]<http://www.htai.org/interest-groups/patient-and-citizen-involvement.html>

A telephone helpline was available for those who were not able to complete the questionnaire online, and in three cases a paper questionnaire was sent and completed by participants and manually entered by the researcher.

The Delphi study was conducted over five months (September 2013 to January 2014). Each round of the Delphi process was open for four weeks, with a two to three week turn around for researchers between each stage. Detailed information about the project was sent to participants two weeks before the Delphi-survey started, to ensure respondents had informed decision making around participating and to optimise response rates in each round of the Delphi. Consent was gained prior to participation in the survey. Reminder emails were sent two weeks after the start of each Delphi survey if experts had not responded. Unique identifiers were used to enable personalised emails containing a survey link to be sent to participants to aid survey administration and to allow monitoring of responses and issue of timely reminders to non-respondents. Confidentiality was maintained and all questionnaires were only identifiable by a code, with all data kept on a file protected computer system. Only amalgamated results were reported rather than individual responses.

i) Round 1

In round 1, participants were asked to rate each of the checklist items from 1 to 10, or no judgement. A rating of 1 meant that the respondent considered the item to be unimportant and should be dropped from checklist. A rating of 10 meant that the respondent considered the item to be very important and must be included. Each point on the scale had a descriptor. Space was provided against each item for free text comments to suggest refinements, reiterations or to suggest additional items.

One hundred and forty-three experts took part in this first round. Two researchers analysed the results of round 1 and free text comments were examined to inform any potential additional items. Consensus was defined by the consistency of median scores (median ≥ 8 = high importance, median 6 or 7 = moderate importance, and median ≤ 5 = low importance), and the absence of significant issues noted in the text comments. Items reaching high importance or moderate importance were selected for round 2 of the Delphi. Free text comments were analysed thematically to identify additional items for round 2.

ii) Round 2

In round 2, participants were asked to rate the GRIPP items again, including any additional items suggested in free text data in round 1. For each item, panellists were given their previous rating, and were also presented with group summary ratings (medians, IQRs), along with all anonymised free text comments from round 1.

One hundred and twenty three experts took part in the second round. Two researchers analysed the results of round 2 and agreed items to be included in the third round. The panellists were asked to re-rate the items and add further comments, if desired. As with the first Delphi round, consensus was defined by the consistency of median scores between rounds (median ≥ 8 = high importance, median 6 or 7 = moderate importance, and median ≤ 5 = low importance), and the absence of significant issues noted in the text comments. If items scored a median score of ≥ 8 in round 1 and round 2, with low inter quartile ranges (IQRs), they had reached consensus for inclusion.

iii) Round 3

For round 3, the results of items reaching consensus (from rounds 1 and 2), and any additional feedback were presented. Round 3 also included new items introduced in Round 2, and items rated of moderate importance (median score 6 or 7) in either or both round 1 and round 2). Additionally, phase 3 included items where comments suggested that single items contained multiple concepts of differing importance. For these, concepts were delineated and respondents were asked to rate each sub item separately. Items scoring ≤ 5 in both round 1 and round 2 were excluded.

The Delphi process is summarised in figure 1.

3. The face-to-face consensus meeting

The final stage of the project was a 1 day consensus meeting with 25 experts representing a range of key stakeholders, including lay representatives (n=8), health and social care organisations (n=6), journal editors (n=2), and academics working in the field of PPI (n=9). The agenda for the day included scene-setting for this event through presentations on relevant background topics, including details of the systematic review evidence and results of the Delphi exercise. Materials were sent to participants 2 weeks before the meeting, including: the agenda, participant list, one or two key papers, and the results of the Delphi exercise. Consent was gained on the day of the meeting.

The face-to face consensus meeting followed an approach similar to the Nominal Group Technique [23,24] by using small group discussion, sharing of ideas and voting techniques. The detailed discussions at the meeting focussed on those items which only reached moderate consensus (n=7). Participants were divided into 4 roundtable groups, with a diverse group of stakeholders in each group. Each group had 20 minutes to discuss each of the items. Participants were encouraged to voice their opinions, with a prerequisite that 'all were equal and every contribution is valid' [25]. Opinions arising during their conversations were captured through the use of different media, for example colour coded cards, post-it notes, and large pieces of paper placed on the tables. This method is intended to enhance creative thinking, expression and communication [25,26]. While a professional facilitator hosted the meeting, each table also nominated a 'table host' whose role it was to keep a focus and to encourage all participants to contribute to the discussion.

After the discussion, each table was asked to feedback their comments on each item. After all tables had fed back, each table had a further 5 minutes to discuss their decision on whether or not to include the item, and one vote from each table was recorded. Consensus on whether to include or exclude the item was achieved if 3 or more tables were in agreement. If two tables voted to include, and two tables voted to exclude, further discussion as a group was conducted, with the facilitator recording each viewpoint on a flip chart using the words spoken by the participants. Consensus was then gained through individual votes.

The second half of the consensus meeting discussed issues arising in respect of content and face validity of the checklist. Participants were asked to check the wording and make any suggested changes to wording directly onto each item. Comments around the comprehensibility of the item were also sought. Suggested modifications were made by the research team following the consensus meeting.

A key session towards the end of the face-to-face meeting was held to discuss the ‘knowledge translation’ strategy to assist with the translation of the reporting guideline into practice. A publication strategy was also developed, and discussions included how the guideline can be implemented by journals.

Evidence of the methods used for developing reporting guidelines have been reported elsewhere [27,28].

DISCUSSION AND CONCLUSION:

This paper has detailed the development of consensus on the items that researchers should consider when reporting PPI. The lessons learned are described below.

Important aspects and lessons learned from the development process

i) Systematic Reviews

While the evidence identified in the systematic review was sufficient to identify key areas of importance to structure specific criteria for the reporting guidelines, several pitfalls can be highlighted in conducting systematic reviews around PPI.

Searching databases to identify potential papers for these reviews provided a number of challenges. As there are no Medical Subject Heading (MeSH) for ‘PPI’ a combination of search terms was used for the electronic databases. A lack of specific search terms led to a lack of sensitivity in the initial searches, resulting in a large number of papers identified initially [1,2]. Databases are not consistent in their indexing of studies relating to PPI which poses many challenges for developing search strategies that aim to locate these papers. Databases also vary the search

terms used, which means the search strategies need to vary by database, increasing the complexity of searching and the potential for error.

This phase of the study was therefore cumbersome and costly. Standardising the terms used for PPI would not only improve search strategies for future reviews of PPI evidence, but also improve comprehension around PPI among health researchers. Standardised terms could then be adopted by electronic databases improving the MeSH search terms. Care was taken with decisions about where and how to search. For example, a dearth of peer-reviewed evidence on PPI studies indicated the importance of searching the grey literature. Restricting the searches to electronic databases, which consist mainly of references to published peer-reviewed journal articles, could have excluded many PPI papers, leading to publication bias.

While it is recommended that all abstracts are reviewed by two researchers in systematic review methodology [11], this greatly increases the cost and time of the process [29]. Evidence shows only an 8% improvement in identification of relevant papers when all abstracts are reviewed by two researchers [11]. A large number of abstracts were retrieved in the searches, therefore for pragmatic reasons, 10% of abstracts were reviewed by two researchers. A high level of consistency was reported, indicating the reliability of the searches.

The quality of the evidence was also very difficult to assess as the PPI evidence was often inconsistently reported and difficult to locate. For example, PPI studies were sometimes reported in full in either the methods or the discussion of the paper [1,2]. Quality assessments were therefore conducted to assess inclusion, but not used to weigh the papers. If studies were fatally flawed in terms of their quality, they were excluded. Care should always be taken when interpreting the results of critical appraisals as they can be biased because of the subjective nature in which decisions are made by researchers. One study compared CASP with the Quality Framework (QF) and intuitive judgement by expert opinion [30] and found that no difference could be detected between the different critical appraisal tools and intuitive judgement by expert opinion.

ii) Delphi Survey

While a three-phase online Delphi survey was chosen to gain consensus on the reporting guideline, other methods were considered including nominal groups [31], analytic hierarchy process (AHP) technique [32] and use of separate working groups gathering consensus through focus groups [33].

From a pragmatic view point, Delphi methods allowed a large number of geographically dispersed stakeholders to be involved in the consensus process which may not have been possible through alternative consensus processes due to time and cost limitations [34]. Respondents could complete the questionnaire at their leisure and this reduced time pressures and may have allowed for more reflection and contemplation of responses [35].

Additionally, the advantage of this method is that members remain anonymous in responding to individual questions and this is likely to encourage opinions that are free of influences from others and more likely to be ‘true’ [36,37]. It also provides an iterative process with controlled feedback, and average score responses from a group of experts, providing stability of responses. Furthermore, it recognises and acknowledges the contribution of each participant. This method can therefore facilitate consensus where there is contradictory or insufficient information to make effective decisions [35-38].

The disadvantages of Delphi methods, which are perhaps true for any consensus method, are that there is a purposeful selection of ‘experts’ chosen because the respondents’ reputations are known to the researcher. These experts meet a minimal number of criteria of familiarity with the research field, self-rating their expertise [39]. Furthermore, the self-selected sample may be biased in that they are willing to take part and therefore either more favourable and more inclined to agree with items included in the GRIPP2 checklist, or disapproving and more likely to disagree with the items included [39].

Delphi methods traditionally use two or more rounds to gain consensus [40]. The number of rounds depends on the level of initial consensus gained, but may also be controlled by time and cost limitations. Studies focussing on the number of rounds needed in a Delphi survey to achieve consensus suggest that most changes occur in

the transition from the first to the second round [41]. When the number of rounds exceeds four, the response rates can be very low due to the response burden on participants [41].

The possibility that participants may alter their estimates in order to conform to the group (conformance), without actually changing their opinions (consensus), was considered. However, evidence suggests that the influence of expert knowledge helps move towards consensus rather than conformance with the 'median score moving towards true value' [42].

An electronic survey was chosen over postal surveys. This allowed a fast deployment of surveys and a quick return time. It was relatively low cost, removing the cost and time of printing, posting and data entry. It also saved the participants from the inconvenience of posting a survey back to us. The disadvantage of this method is that the participants had to have access to a computer and be computer literate. This may have biased the sample, particularly excluding some hard to reach patients and carers.

An online survey software, Snap Chat, was adapted by in-house expertise to record the survey responses [21]. Other electronic packages were considered such as Survey Monkey and Zoomerang, GoogleForms, and SurveyGizmo [43-46], but Snap Chat was internally available and proved a reliable and flexible survey software for the study's requirements. Furthermore, using in-house expertise allowed us to tailor and customise the online form to our needs more effectively thus allowing us to refine the design and functionality after piloting. Survey responses were monitored and the survey data were cleaned and analysed to produce the results in the required format.

The electronic software offered a save and return function within the survey, allowing participants to reflect and return to the survey. This also would have minimised the risk of incomplete responses where participants sometimes underestimate the time the survey will take and then run out of time to complete the questionnaire in full.

Consideration was given to the number of categories in the Likert scale for rating the items within the GRIPP2 Delphi survey. Streiner and Norman (2008) have argued that the benefit from large numbers of options is subject to the law of diminishing returns [47], and that from the 7-item scale and upwards, the scales become too cumbersome to use. Any additional benefits are cancelled out by respondent fatigue and reliability plummets [47]. However, in reality we are often asked to rate issues on a scale of 1 to 10, it provides a better opportunity to detect and discriminate when responses are skewed at one end of the scale, and it felt more ‘natural’ to patient advisors on the research team [48]. This is also how previous Delphi surveys had been scored in the development of reporting guidelines through EQUATOR.

Different descriptive statistics can be used when feeding back data in each round. While the mean, mode and median scores provide three forms of averages [49], the range, inter-quartile range (IQR) and standard deviation are all measures that indicate the amount of variability within a dataset [49]. As there was the potential for a range of scores for each item, the median was selected as a better score of the average than the mean or the mode. The range is the simplest measure of variability to calculate but can be misleading if the dataset contains extreme values. The IQR reduces this problem by considering the variability within the middle 50% of the dataset. The standard deviation is used to take into account a measure of how every value in the dataset varies from the mean. The interquartile range was therefore used to report the variation from the median scores [50].

The validity and reliability of a Delphi study may be questioned due to the subjective nature of the feedback and potential instability of responses. However, this approach engages a wide range of expertise more effectively than any other group consensus method, and provides a ‘fair’ representation of the views of each participant because each participant has an equal opportunity to have their views taken into account. Furthermore, Delphi methods clearly state the rationale for inclusion or exclusion in the final checklist, whereas other methods of gaining consensus rarely provide such a transparent decision trail for each item. The quality of the Delphi survey was further increased by the quick turnaround between Delphi rounds [51]. The Delphi study for the reporting guidelines provided opinions from a

representative sample of all PPI stakeholders, carefully and rigorously collated over 3 rounds of voting. In this study the Delphi methods used brought agreement from a diverse group of stakeholders whose commitment to the project resulted in good response rates at each stage of the Delphi project. The validity and reliability of the process was therefore deemed satisfactory.

Finally, ethical issues were also examined in this Delphi survey. The main ethical issues related to consent to participate, privacy, and confidentiality of the data provided, all of which were addressed in the methods used.

iii) Consensus Face to Face Meeting

The consensus face-to face meeting was important to finalise consensus on the reporting guideline. An informal approach to voting was adopted where consensus was gained through the round robin process with a gradual move towards synthesis and building consensus rather than through anonymous voting [51]. This approach encourages the sharing and discussion of the reasons for the choices made by each group member, thereby identifying common ground, and a plurality of ideas and approaches [51]. The meeting also ensured face and content validity through small group discussion.

The success of the consensus group was enhanced by careful planning and commitment of the stakeholders. Furthermore, a critical role in the success of the consensus group was that of the facilitator. Key aspects of this role were ensuring that participants of the group understand their roles, and adopting the listening stance so participants all felt listened to. This important role helped to ensure that balanced views were recorded through individual and group work. Facilitated groups develop greater consensus than user-driven groups [52,53]. However, highly structured facilitation can have an adverse effect on the consensus process, and element of flexibility in the process is recommended [52,53]. A highly skilled facilitator was used to mediate the group process, and to ensure key and timely contributions from all members. The facilitator presented items where consensus decisions were needed, guided the participants to reach agreement, ensuring all have an opportunity to participate. For the items related to economic assessment and testing

conceptual or theoretical models, where decisions on whether to include or not were difficult, the facilitator used problem solving techniques to finalise the decision. This involved the group drawing up a list of the pro’s and con’s of the item, and asking them to review and evaluate the list and then to re-evaluate their initial decision.

An important aspect of the consensus meeting was also agreeing a plan of action for dissemination, implementation and adoption of GRIPP2. Evaluation of GRIPP2 will be conducted through ongoing feedback from authors.

Table 1 below summarises the lessons learnt during this study:

Table 1: Summary of lessons learnt:

Systematic Review	<ul style="list-style-type: none"> • Lack of MESH terms for PPI in databases made searches broad • No consistency in indexing between databases lead to development of separate search strategies for each database • Difficulty identifying PPI data as written up in methods or discussion sections of papers • Pragmatics of all titles reviewed by two researchers led to 10% reviewed twice
Delphi Study	<p><u>Advantages:</u></p> <ul style="list-style-type: none"> • Pragmatic consensus approach to gain large, geographically dispersed group • Participants complete the questionnaire at their leisure • Participants remain anonymous so greater openness and honesty of comments <p><u>Challenges:</u></p> <ul style="list-style-type: none"> • Selection bias – as self-selected. Tendency to recruit participants who are more favourable or more disapproving. • Excludes those who do not have a computer and those who are not computer literate. • Decision-making on number of Delphi rounds to gain consensus. Need enough to gain consensus but not too many that you have low response rates for subsequent rounds. • Difficulty of choosing off the shelf data collection packages – importance of customising software. • Issues of response fatigue with too many categories in the Likert scale – the higher the number the greater the response fatigue. • Need for clear instructions, bibliography, and other information in lay language. • Use of median score and inter-quartile ranges vs mean (sd) and modes in presentation of data – to overcome extreme values in dataset
Consensus Meeting	<ul style="list-style-type: none"> • Need for openness and informal approach to encourage discussion and plurality of ideas. • Critical to have expert facilitator to guide the consensus meeting to ensure key and timely contributions from all members • Importance of planning for dissemination, implementation and adoption of outcomes

Contribution of the patient partners

The patient partners contributed to the development of GRIPP in a number of ways. Throughout the initial stages of collating the evidence and identifying items for the GRIPP checklist, the patient partners highlighted the importance of including items referring to the context and processes of PPI, suggesting that this affected the impact that PPI had on research. The patient partners, along with other patient organisations and charities, recruited nearly half of all participants for the Delphi survey. The patient partners helped other patients with the technical aspects of completing the online survey, improving the response rate in each Delphi survey round. The patient partners not only checked the comprehension of the changed items and comments from the lay perspective between rounds, but were also integral to helping the researchers keep to the scheduled time of the Delphi survey. Throughout the write up phase for both the results paper and the methods paper, the patient partners contributed to the lay sections and contributed to edits of the paper.

Conclusion

In conclusion, this paper details the methods used in the development of EQUATOR recognised guidelines for the reporting of patient and public involvement (PPI) in research (GRIPP2) [9].

GRIPP2 has been developed using the robust methods used in the development of other EQUATOR guidelines such as CONSORT and STROBE. The development process involved identification of relevant evidence through systematic reviews, consensus of included items through an online Delphi survey with PPI experts, and a face to face meeting of PPI experts to finalise consensus.

Systematic reviews informed the development of items for the reporting guideline. However, lack of MESH terms for PPI, inconsistency of indexing between databases, large number of titles in searches, and difficulty locating evidence of PPI in the papers was time consuming. An online Delphi survey provided a pragmatic and anonymous process for consensus, although challenges were encountered with selection bias in the sample, avoiding response fatigue and decision-making over the

1
2
3 presentation of data. The success of the consensus meeting was due to careful
4 planning and the critical role of the facilitator.
5
6
7

8
9 GRIPP2 has been developed to improve the accuracy and consistency of PPI reporting in
10 health research to improve interpretation and better application in future research. With a
11 growing evidence base, we expect greater discussion around the conceptualisation and
12 theoretical underpinning of PPI to provide a greater insight into practices and processes
13 [54,55,56] As the field of patient and public involvement develops we expect further
14 development of GRIPP2 to refine and update the guidance.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References:

1. Brett J, Staniszewska S, Mockford C, et al.. Mapping the Impact of Patient and Public Involvement (PPI) on Health and Social Care Research: A systematic review. *Health Expectations* 2014;17(5):637–650

2. Mockford C, Staniszewska S, Griffiths F, and Herron-Marx S (2011). The impact of patient and public involvement (PPI) on UK NHS healthcare: a systematic review. *International Journal for Quality in Health Care* 2011; 24(1):28-38

3. Staniszewska S, Brett J, Mockford C, Barber R. The GRIPP checklist: Strengthening the quality of patient and public involvement reporting in research. *International Journal of Health Technology Assessment in Health Care* 2011; 27 (4): 391-399

4. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010 Mar 23;340:c332. doi: 10.1136/bmj.c332.

5. von Elm E, Altman DG, Egger M, et al.; STROBE Initiative The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007 Oct 20;370(9596):1453-7.

6. Moher D, Altman D, Shultz K, Simera I, Wager E. Guidelines for Reporting Health Research: A User's Manual. *BMJ Books*. Sept 2014

7. Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med* 2010;8(1):24. PMID: 20420659

8. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med* 2010; 7(2):e1000217. PMID: 20169112

9. Staniszewska S, Brett J, Simera I, et al. Guidance for reporting patient and public involvement in health research: A Delphi consensus study to develop GRIPP2-SF and GRIPP2-LF. *BMJ Accepted* June 2016

10. Cook DJ, Mulrow CD, Haynes RB. Systematic Reviews: Synthesis of Best Evidence for Clinical Decisions. *Annals of Internal Medicine*. 1997;126(5):376- 80

11. Centre for Reviews and Dissemination (CRD). Systematic Reviews: Guidance for undertaking reviews in health care. University of York, 2008 <https://www.york.ac.uk/>

12. INVOLVE: InvoNET website www.invo.org.uk. Accessed June 2012.

13. NHS Evidence www.evidence.nhs.uk. Accessed June 2012

14. Kings fund Library. www.kingsfund.org.uk/library. Accessed June 2012
15. National library for Health www.library.nhs.uk. Accessed 2012
16. Joseph Rowntree Fund Library . www.jrf.org.uk/Library. Accessed June 2012
17. Critical Appraisal Skills Programme (CASP) 2014. CASP Checklists (URL used) Oxford. CASP
18. Dixon-Woods M, Kirk D, Agarwal S et al. Vulnerable Groups and Access, to Health Care: A Critical Interpretative Review. London: National Co-ordinating Centre for NHS Service, 2005.
19. Hubbard G, Kidd L, Donaghy E, McDonald C, Kearney N. A review of literature about involving people affected by cancer in research, policy and planning and practice. Patient Education and Counseling, 2007; 65: 21–33.
20. Williams PL, and Webb C. 'The Delphi technique: a methodological discussion', Journal of Advanced Nursing 1994;19: 180-186.
21. Tetzlaff J, Moher D, Chan A. Developing a guideline for clinical trial protocol content: Delphi consensus survey. Trials 2012;13: 176
22. Snap chat www.snapchat.com. Accessed Sept.2011
23. Harvey N, Holmes CA: Nominal group technique: An effective method for obtaining group consensus. Int J Nurs Pract 2012;18:188-194.
24. Gallagher M, Hares T, Spencer J, et al: The nominal group technique: A research tool for general practice? Fam Pract 1993;10:76-81.
25. Brown JID.. The World Cafe. Shaping Our Futures Through Conversations That Matter (p. 765). Berrett-Koehler Publishers. Inc 2005.
26. Wilkinson JE, Rycroft-Malone J, McCormack B. A creative approach to the development of an agenda for knowledge utilization: outputs from the 11th international knowledge utilization colloquium (KU 11). Worldviews on Evidence Based Nursing 2012; 9(4): 195–199.
27. Moher D, Schulz KF, Altman DG. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 2001;357:1191- 1194.

28. Moher D. Guidelines for reporting health care research: advancing the clarity and transparency of scientific reporting. *Can J Anaesth.* 2009;56:96-101

29. Attree P. Growing up in disadvantage: a systematic review of the qualitative evidence. *Child Care Health Dev* 2004;30:679-89.

30. Dixon-Woods M, Sutton AJ, Shaw RL, Miller T, Smith JA, Young B, Bonas S, Booth A, Jones DR. Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods. *Journal of Health Services Research and Policy* 2007; 12(1): 42-47.

31. Carney, O., McIntosh, J. and Worth, A. (1996) 'The use of the nominal group technique in research with community nurses', *Journal of Advanced Nursing* 1996; 23 (5):1024-1029.

32. Lai VS, Wong BK, Cehung W. 'Group decision making in a multiple criteria environment: A case using the AHP in software selection', *European Journal of Operational Research* 2002; 137: 134-144.

33. Morgan DL. *Focus Groups as Qualitative Research. Planning and design of focus groups* 2nd Edition. SAGE Research Methods London 1997

34. Rogers MR, and Lopez EC. Identifying critical cross-cultural school psychology competencies. *Journal of School Psychology* 2002;40:115-141.

35. Linstone HA. and Turoff M. *The Delphi Method Techniques and method?* *Nursing Research* 1975; 46(2): 116-118.

36. Goodman CM *The Delphi technique: a critique.* *Journal of Advanced Nursing* 1987;12: 729-734.

37. Snyder-Halpern R. Indicators of organizational readiness for clinical information technology/systems innovation: a Delphi study. *International Journal of Medical Informatics* 2002;63: 179-204.

38. Hasson F, Keeney S, McKenna H. (2000) Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing* 2000; 32(4):1008-1015.

39. Keeney S, Hasson F, McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. *Int J Nurs Stud.* 2001 Apr; 38(2):195-200.

40. Cuhls K. *The Delphi Method.* Fraunhofer Institute for Systems and Innovation Research, Germany 1998

41. vanZolingen SJ. and Klaassen CA. Selection processes in a Delphi study about key qualifications in Senior Secondary Vocational Education, Technological Forecasting and Social Change 2003; 70: 317-340.

42. Munier F, Rondé P. The role of knowledge codification in the emergence of consensus under uncertainty: empirical analysis and policy implications. Research Policy 2001;20: 1537-1551 .

43. Survey Monkey <https://www.surveymonkey.com/> 1999-2016

44. Zoomerang (online web survey) <http://www.zoomerang.com/> 1999-2016

45. Google forms (Online web survey) www.google.com/forms February 2007

46. Survey Gizmo <https://www.surveygizmo.com/> 2005-2016

47. Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use 2008

48. Wittnk DR, Bayer LR.. "The Measurement Imperative" Marketing Research 2003;6(4): 14-22

49. Altman D. Practical Statistics for Medical Research (Chapman & Hall). November 22, 1990

50. Murphy MK., Black NA., Lamping DL., McKee CM., Sanderson CDB., Askham J, Marteau T. Consensus development methods, and their use in clinical guideline development: a review United Kingdom, Health Technology Assessment 1998; 2(3):49

51. Dunham RB. Organisational Behaviour: The Delphi Technique, University of Wisconsin School of Business, US 1998

52. Anson R, Bostrom R, Wynne B. An Experiment Assessing Group Support System and Facilitator Effects on Meeting Outcomes. Management Science 1995; 41:2

53. Dickson G, Partridge J, Robinson L. Exploring modes of facilitative support for GDSS Technology MIS Quarterly 1993; 17(2):173-194

54. Ives J, Damery S, Redwod S. "PPI, paradoxes and Plato: who's sailing the ship?." *Journal of Medical Ethics* 2013; 39: 186-187

55. Brett J, Staniszewska S, Mockford C, et al. The PIRICOM Study: A systematic review of the conceptualisation, measurement, impact and outcomes of patients and public involvement in health and social care research. UK-CRC 2010

56. Forbat, Liz, Gill Hubbard, and Nora Kearney. "Patient and public involvement: models and muddles." *Journal of clinical nursing* 18.18 (2009): 2547-2554.

For peer review only

Statement of contribution

Staniszewska, Brett, Simera, Seers, Mockford, Goodlad, Morris, Denegri, Barber, Littejohn, Thomas, Suleman, Entwistle and Tysall made substantial contributions to conception and design. Brett, Staniszewska, Simera, Altman, and Moher made substantial contributions to developing the protocol. Staniszewska, Brett, Goodlad made substantial contributions to the acquisition of data, analysis, and interpretation of data. All authors have been involved in drafting the manuscript or revising it critically for important intellectual content; given final approval of the version to be published.

Transparency declaration: the manuscript is an honest, accurate, and transparent account of the methods for the study being reported; no important aspects of the study have been omitted; and there are no discrepancies from the study

Acknowledgments

We are very grateful to everyone who participated in the Delphi survey and attending the consensus event. GRIPP 2 was funded by the RCN Research Institute, Warwick Medical School, as part of its strategic focus on patient and public involvement. We are grateful for Sally Crowe who facilitated the consensus workshop. SS is part funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care West Midlands. This paper presents independent research and the views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. PL is supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South London at King's College Hospital NHS Foundation Trust. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health

Compliance with Ethical Standards:

Ethical approval for the study was secured by the Centre for Education Development, Appraisal and Research (CEDAR) at the University of Warwick who secured permission to carry out the Delphi Survey from the University of Warwick BSREC committee

Ethical approval for the study was secured by the Centre for Education and Industry. Appraisal and Research (CEI) at the University of Warwick who gained generic approval from the University Ethics Committee for all of its online surveys. The committee reviewed the rigorous survey procedures CEI had in place and granted

generic ethical approval for its robust process and procedures. The GRIPP2 Delphi survey was assessed by CEI as being covered by the generic approval.

Funding: This study was funded by RCNRI, Warwick Medical School, University of Warwick

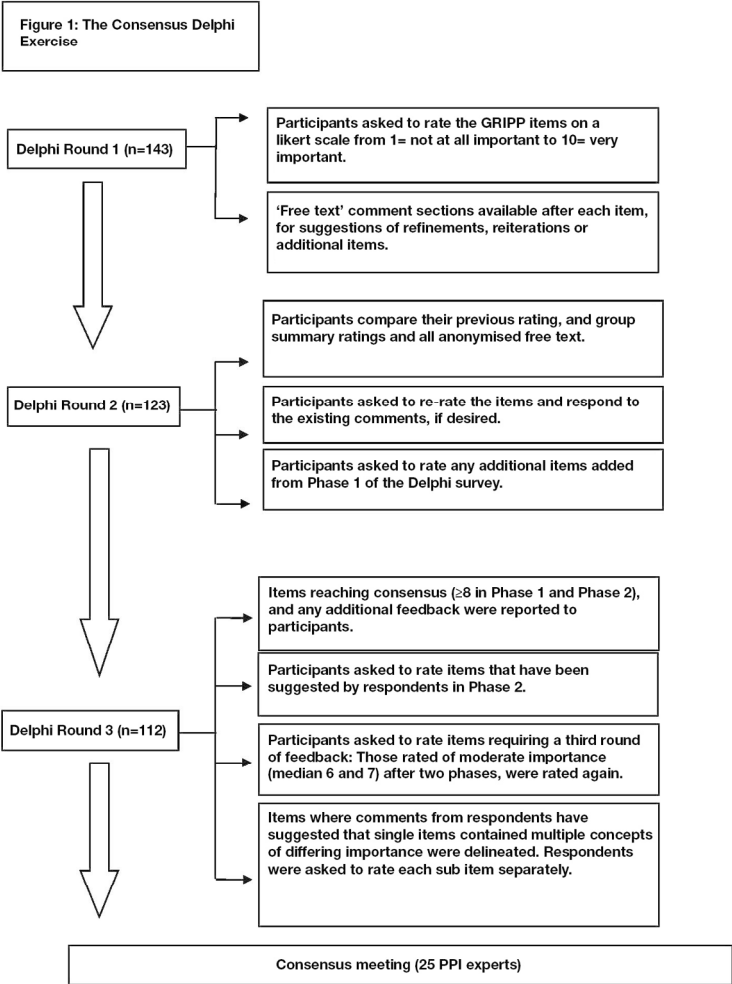
Data-Sharing: De-identified data will be shared through university accessible databases or repositories at Warwick University and Oxford Brookes University. Please contact Dr Jo Brett if additional information is required: Email: Jbrett@brookes.ac.uk

Conflict of Interest: No authors reported conflict of interests

Figure 1: The Consensus Delphi Exercise

For peer review only

Brett J. GRIPP2 Methods 031717 report bmjopen-2017-016948



210x297mm (200 x 200 DPI)